JPRS-UCC-88-003
4 OCTOBER 1988

FOREIGN BROADCAST INFORMATION SERVICE

JPRS Report

Science & Technology

USSR: Computers

JPRS-UCC-88-003

4 OCTOBER 1988

SCIENCE & TECHNOLOGY

USSR: COMPUTERS

CONTENTS

/9835

UDC 681.325

Changes in High-Speed Computer Systems

18630203a Kiev UPRAVLYAYUSHCHIYE SISTEMY I MASHINY in Russian No 6, Nov-Dec 87 pp 39-46

[Article by S. B. Pogrebinsky: "Developments in the Architecture of High-Speed Computers]

[Text] Introduction. In recent years, there have been several different trends in the development of the architecture of high-speed computers designed for use in large computer centers, automated design systems, systems for economic modelling and management and for building educational and testing systems.

The goals of the systems under development may sound somewhat traditional:

--greater capacity;
--improved reliability;
--and greater intelligence of user-interactive software.

Greater computer capacity is needed to study objects and phenomena that are characterized by large geometric dimensions, complex shapes and extended time duration. The study of such objects and phenomena has increased the need for modelling precision, i.e., it demands a more discrete and continuous process, which in turn substantially increases the amount of data handled.

There have been qualitative and quantitative changes in the problems of automating design work and in achieving optimal control over and study of operations. More and more often now, to reach optimal solutions, it is necessary to examine complex systems in a dynamic state and to study the different ways in which these systems behave.

In addition to purely computational problems, the processing of complex data structures, controlling large data bases and knowledge-bases, expert systems, logic systems and computer scanning and input of graphic material have all gained greater importance. All of these problems demand greater computer speed. The need to improve reliability has to a large extent been dictated by the greater practical importance placed on the results of computer calculations and the correspondingly greater cost of errors. This is true not only of control systems, but also of those used for design and research computations.

A mistake (or failure to obtain a result) can be the result of a failure of any one of the three basic components of a computer system: hardware, system software or applications software.

The poor performance of hardware may be the result of previously-undetected errors in its design or manufacture or the failure of a component.

Poor program performance is often caused by undetected design errors or incorrect use. The growing complexity of hardware and software makes it highly probable that many other design errors are still undetected.

A study (reference 1) of system failures caused by failures and errors in apparatus or programs shows that, regardless of the differences in their physical nature, general tendencies in failures of both hard- and software are much the same and can be examined according to general methods of reliability theory.

Not only is it necessary to perfect the equipment and technology used to design and program computers, but it is also necessary to change the architecture of computers to enhance the reliability of their hardware and their operating systems and applied programming.

Let us note that among the causes of poor software reliability are poorly-developed user-computer communications and the substantial gap between programming languages and computer architecture--the informational objects and operations undertaken by the computer do not, as a rule, coincide with objects and operations expressed in programming languages.

Improved intelligence of communications software and hardware, the use of user-friendly concepts and objects (graphic images, acoustics, general mathematics and subjects and operations using them, hierarchical data structures, frames, means of associative retrieval and the study of object properties) also require the elimination of this great discrepancy between the languages of the user and systems programmer versus the language of the computer. A considerable increase in speed is also necessary to ensure acceptable reaction time in computer-user dialogue and in control of the computer system.

Multiprocessor architectures. A decisive direction in the development of a computer architecture aimed at achieving high speed and reliability has been the ability of several types of computers to simultaneously perform various parts of one or several commands and large blocks of a program. Unfortunately, such possibilities are, for the most part, highly limited; and it has not been possible to attain any substantial success in the effort.

The search for ways of increasing computer speed through the simultaneous performance of various elements of a program is taking several different turns.

First of all, problem-solving algorithms in which such a possibility is more or less apparent are being studied. For example, there are problems in which

the same simple operation is performed on a large amount of homogeneous data. This includes many of the tasks of mathematical physics, data-searching, observed-data processing and linear algebra, etc. To solve such problems, it seemed enough to build efficient computers wirh high speed and economy of operation. Among them, the two most widespread types of architectures are the pipeline computer and process processors with single-command-flow and multiple-data-flow architecture.

Second, many design and data-processing tasks may characteristically by subdivided into relatively independent program blocks within which shared data is processed. Such a subdivision can be carried out during the creation of both the physical and mathematical model. It is also possible that several independent program relatively using the same data field can be run simultaneously. Such an approach led to the creation of a computer with parts multiple-command-flow and multiple-data-flow (MKMD) architecture. According to the definition, a computer with this type of architecture uses several identical data-processing devices (with, perhaps, a local memory) that are linked to a common memory device and work under an integrated operating system.

The second approach was to search for new ways of organizing data and algorithms and new principles for organizing the computing process in such a way as to reveal the natural parallels that exist among most problems and express those parallels in algorithms.

This approach took two directions.

One well-known implementation of such an approach to the design of computer architecture is the principle of checking the computation of data flows. In contrast to traditional architectures in which the order of program operations is determined by given sequence of numbered commands, when the data flow is checked at every step all commands for which the necessary data is present are performed. In this approach command order, the results of every operation performed are sent immediately to all program commands that will use them.

Several designs using this approach have been drawn up and experimental studies have revealed substantial difficulties in the realization of multipurpose computers based on data-flow control.

Completely original results have been achieved in implementing the principle of macropipeline organization of computations (references 2, 3). This principle is based on analysis of data structures and their data messages in the mathematical model of the process employed. The model's algorithm is represented as a network of program modules and their data fields, with the maximum number of communications links closed within each module and therefore a minimum exchanges between modules.

Several broad concepts and the corresponding data structures have been formalized for the subject area and general mathematical description of the

problem. These concepts describe the problem to be solved. With such a presentation of the program, the data used by the commands is not single words or bytes but complex data structures. Such a description (programs and data) supported by the computer's resources (interpreters of corresponding high-level languages) makes it possible to plan task performance efficiently and break it down into large program blocks that are largely independent of each other. The order of simultaneous performance of program blocks (their initiation, synchronization and data exchange) is maintained by the concepts of the high-level language; it is thus possible to check tasks, data and equipment in parallel.

The use of the principle of macropipeline organization of the computation process has made it possible to develop an architecture and create a multiprocessor computing complex with several hundred processors in which a linear increase in capacity is practically assured as the number of processors used is increased.

Such a computer complex is distinguished by a high degree of parallelism and adaptability. It can create an adequate functional analogue of the physical situation under study in which the simultaneously and independently performed processes used to solve the problem are directly reflected in the composition of hardware and dynamically limited communications structures.

The creation of such an adequate functional model has been to a considerable degree determined by the success at adaptation, that is the reworking of the architecture of processors by means of "soft" microprogramming and a dynamic change in the communications links among these processors.

It is hard to make an accurate comparison of the various architectures of multiprocessor computers by using the criteria of potential output since they are based on hardware with such a wide variety in size and complexity of processors and capacities and parameters of communications systems as adapted to the given architecture. There is also a reverse relationship in the design of a computer, system of hardware available determines what architecture will be selected.

The possibility of an accurate comparison is made even more remote by the fact that in high-speed computers with the new architecture, it is the software that is the dominant factor in both the speed and cost of the system.

The quality of the operating and programming systems and how well they can be adapted to the selected architecture can enhance or eliminate the inherent advantages and disadvantages of the given computer.

In evaluating the actual capacity of a computer, one must always consider reliability. The user is naturally interested in the time the program takes to solve a problem, but he is even more in the probability that problem will be solved within the expected amount of time. As the program's running time approaches the computer system's mean time between failures, the probability that a solution will be produced is rather small. At the same time, there is a much greater probability that incorrect solutions will be produced.

Increasing capacity within single-process or pipeline computers using internal-command parallelism obviously requires an increase in the amount of sequentially-connected apparatus (i.e., connected in such a way that the failure of one element causes the entire system to fail). This arrangement leads to an increase in system speed and a drop in reliability.

In a computer with MKMD architecture, an increase in the number of processors considerably increases the need for high speed and basic resources (overall operating memory and operating system) and elimination or reduction of the effects of conflicting situations.

Furthermore, this architecture does not solve the problem of localizing rapid exchanges. The speed of every processor of such a computer is defined as the speed of exchange with the shared memory (with an average of two exchanges for every command). The relay providing the exchange between the processors and the main memory should have a very high carrying capacity. For computers with potential capacities of 10^8 operations per second, the relay's carrying capacity should be about 2×10^8 words per second.

This means that computers with MKMD architecture must be very well equipped with basic resources, the failure of which causes the system as a whole to fail. The necessary levels of reliability is not attainable under these conditions.

Distributed MKMD architectures. High reliability and speed can be combined in multicomputer systems that provide distributed performance of a multiple stream of commands over a multiple stream of data.

The distributed MKMD architecture makes possible to parallel performance of all data processing functions important to running the program. Under these conditions there is a linear growth in capacity as the number of processors rises.

In a distributed MKMD architecture, the following operations are performed in parallel;

--control over performance of the required task;
--control of equipment;
--strictly computational and logical operations with data;
--the retrieval and structuring of necessary data;
--exchange of data among various components, such as processors, various levels of memory and input-output devices;
--relaying.

Parallelism is implemented on two levels: all of the listed processes are performed in parallel and each of them has also been made parallel.

The distributed MKMD architecture makes possible a high level of reliability through dynamic creation of configurations. The new configurations (set of processors and links between them) are created during the solution of the

problem as each new step is performed, equipment is freed up, and as the general composition of equipment changes.

We will now discuss several properties of a multiprocessor computer with distributed MKMD architecture.

Autonomy and functional completeness of each processor. Each of the computer's processors is completely independent to run a program, control all types of interrupt (of programs, input-output, internal operations, communications with the supervisor and program control systems) and also cooperate with other participants in the problem-solving process if the program incorporates it (to receive and issue messages and data). The processor performs all of these actions independently of the other participants and is connected to the communications network only in exchange operations.

The distribution of system hardware and software among its components. All apparatus and programs providing input, output and stroage of data, exchange between processors and equipment (main memory ↔ arithmetic logic unit), data processing and process control are divided among the processors. A hierarchically-organized distributive operating system whose program runs in the computer's processors (including those specially created for that purpose) controls task execution and equipment functioning.

General-purpose communications network. The flexibility and adaptability of the multiprocessor computer are to a large extent dependent on the degree of universality of the communications network, i.e., on the ability of all the computer's components to establish links among themselves in pairs and/or groups. As has already been noted, the macropipeline organization of the computing process considerably lowers the need for rapid exchange among elements involved in task completion.

Modelling and experience in problem-solving on the macropipeline computer have shown that a balance is achieved if an approximate parity can be maintained between the average time it taks to perform an operation or in the processor and the time it takes to transmit one byte of data between processors.

Thus, at a processor speed on the order of one million operations per second, data transmission on each channel of the communications network should amount to approximately one megabyte per second. For strongly-linked systems, the relative speed of exchange should be an order higher at least. The complexity and cost of an all-purpose network in communications with simple protocoles is manageable and requires no appreciable increase in the amount of equipment.

Process control with instructions and messages in a high-level language. Through the arrangement of hardware and software, every processor in a multiprocessor computer forms an interpreter of the language of instructions and high-level messages. The use of a high-level language in instructions and messages makes it possible to lower substantially the volume of control information transmitted and to specialize the operating system's control processes. Interpretation of program control organization of the interactions of programs substantially simplifies the planning of computations, the choice

of strategies, and the preparation of data; it also makes possible macro-level control over operations with the data flow.

Organization of user interaction with the computer is also greatly simplified. It is now possible for the user to be continuously informed about the course of the computer's work and for the user to intervene in situations unanticipated by the program.

Interpretation of the control language makes it possible to put off linking programs and equipment until just before the program is run. In a multi-processor computer,,it is the process (set of interacting programs and data comprising the complete problem) and the processor (specific piece of hard-ware in which this problem is to be solved) that stand out. A linkup occurs whenever a process is ready to be performed and a processor is free at the same time.

All of the listed attributes of architecture make it possible to create configurations dynamically in a multiprocessor computer.

There are two kins of configruations: static and dynamic.

In the former, the resources are allotted to solving the problem before the computer begins work. Likewise, before the start of calculation's, the linking mathematical and physical addresses is carried out. In some circum-stances, the components of the multiprocessor system are also joined. If any one of the processors fails and is switched off from the others, it is necessary to redistribute tasks among the reduced number of processors or replace the failed unit if the system has a free one available. In the first case, the plan for solving the problem should be written and references to all program blocks should be re-examined. If the processor is replaced, only the references need be changed. This is fairly complicated because all of the program text's references to the replaced processor must be located.

In dynamic distribution of resources (dynamic reconfiguration), all of the free processors can be selected for a task at any given moment. All exchanges between processors are described in program texts in terms of mathematical addresses alone. Linkage of addresses is done at the moment that the processor turns to the operating system to be loaded and input-output information is requested.

Configuration in a macropipeline computer is a logical rather than a physical concept. The task of configuration is to transmit from the control program to the operating system a list of processes ready to be performed at any given instant of time. As many such lists (configurations) can exist in a multi-processor computer as there are different problems being solved at one time.

At any moment, the operating system can attach the ready process to one of the free processors named in the list of free processors. This attachment consists of establishing a correspondence between the name of the process and

the physical number of the processor in which the program and data forming the process are loaded.

An important feature of a multiprocessor computer is its ability to make dynamic changes in the list of free processors by adding newly-repaired units or disconnecting unneeded or faulty ones (reconfiguration). The list also includes processors that can be used to implement a task control program or the programs of the operating system.

The characteristics of a macropipeline computer's architecture and especially the principle of dynamic creation of configurationsshave made it possible to implement a failure-free "ideology" (failure being the gradual deterioration of parameters of the computing system as its components fail) and therefore considerably increase its reliability. A series of measures have guaranteed the efficiency of the reliability-maintenance system.

First of all, there is the already-mentioned ability to reconfigure the basic equipment, i.e., dynamic disconnection of faulty processors and devices and introduction of repaired components.

The second measure is to minimize the amount of backup and required equipment, i.e., equipment whose failure will make the system impossible. Each kind of processor is used in large numbers, so a few extra of each serves as an adequate backup without significantly increasing the cost of equipment. The network of connections has been designed according to a block-hierarchic principle. A large part of the network equipment is allotted to the processors. We must note that under ideal conditions, all equipment of the communications network would be allotted to the processors.

The third measure is the creation of a complete system of early failure detection in the computer's basic processors. This system includes equipment for on-line apparatus monitoring, programs and microprograms:

--monitoring of all types of memory and data transmission busses;
--microdiagnostics in the microprograms for monitoring processor equipment;
--periodic programmed monitoring of the components of the multiprocessor computer.
--monitoring of the operating system (time checks, transmission checks, etc.).

Every processor has a predictable level of reliability which is determined by software within the operating system that gathers and processes statistical data about equipment failures.

Hardware support of multiprocessor computers. Several properties stand out as the most important factors in determining the efficiency of a multiprocessor computer. These properties are manifested in the composition and structure of the hardware of the computer. They include the following:

--enhancement of program qulaity and efficiency;
--dynamic planning and control of computing processes;

--dynamic creation of working configurations;
--organization of the interaction of processes and processors;
--support for the work capacity of hardware, early detection of failures and localization of their effects.

The structure of the hardware and the internal process languages of a multiprocessor computer should be oriented towards effective implementation of the listed properties.

The difficulties encountered in this endeavor are well known. They are basically the result of the considerable gap between the system of objects and concepts used by people to describe the algorithm and the set of hardware. In multiprocessor computers, the significance of this gap naturally increases with a rise in the complexity of the problems being solved and the complexity of the organization of the computing process (reference 4).

Essentially, there are two different problems encountered in the design of multiprocessor computers. The first of these involves the orientation of the architecture toward building programming systems and running those programs on the basis of high-level algorithmic languages and the second involves the orientation of the architecture for support of the organization of multiprocessor computations.

In both cases, there is an obvious shortcoming in that the computer language uses only the simplest data elements such as codes, numbers and symbols and their corresponding memory organization.

The computer language should include a hierarchic system of objects with operations for recognizing their properties and transformations, and in terms of which the user can describe the problem. Every object in such a system is represented by an entire set of informationally and topologically interrelated elements of data and by a set of operations used to handle them. One can define several properties of such objects.

<u>Completeness</u>. All elements of the object can only exist jointly. They are created and destroyed at the same time. Operations involving an object affect the entire object. An object is addressed as a unit. If required, there does exist a rule that defines the means of computing the addresses of elements within an object.

Operations are defined to shift objects as a whole, combine or create structures from them or divide them into independent parts, each of which becomes an independent object.

<u>Self-identifiability</u>. The object includes its own identifying data, which defines its individuality, properties and status at any given moment.

<u>Functionality</u>. The object is designated to process ceratin data--to recognize its properties, store and move it.

In accordance with the concept of an object, in high-speed computers, there are also changes in the methods of orgainizing memory, which is dynamically broken down into areas of object location. As a rule, a local high-speed memory calls up an object in its entirety. Let us note that such a method of addressing fits in well with layering memory to increase its speed since the elements of the object can be located in neighboring elements of the layered memory and all be called up at once.

We should also look at other methods of memory organization that correspond to the structures of objects (such as stack, serial, etc.).

The defining element of a macropipeline computer is the processor. The concept of a process in a multiprocessor computer differs slightly from the traditional concept. As we have already said, in a multiprocessor computer, all apparatus and programs providing input-output, storage and data-processing and control over the computation process are distributed among a number of processors.

Each of these includes a data subprocessor (which corresponds to the traditional concept of the central processor), channel subprocessor and working memory, all interconnected by a high-speed communications port.

The multiprocessor computer uses several types of processors:

--an arithmetic processor for processing numerical data;
--a control processor for processing symbolic data and fulfilling control tasks;
--specialized processors;
--service processor.

The use of processors in a macropipeline computer demanded a specific architectural orientation. This includes the following.

The ratio of processor parameters (speed : memory size : input-output carrying capacity : carrying capacity of the communications port) is selected to maximize turnover in terms of input data (number of operations per byte of data introduced). This makes it possible to design a problem-solving program in which each processor receives a substantial, autonomous task. Experience in solving problems on a macropipeline computer and modelling have shown that a parameter ratio of 1:1:2:32 (i.e., one million operations/second : 1 megabyte of memory : 2 megabytes/second : 32 megabytes/second) is sufficient for the arithmetic process.

The structure of the processor includes hardware support for the functioning of the control program itself and other parts of the operating system.

Certain operating system support apparatus is intended to organize sequences, handle interrupt, etc. It is necessary to point out that each individual processor works only on a single program, since in a multiprocessor, multiprogramming is carried out by multiple processors.

The internal languages of the processors in multiprocessor computers to a large extent reflect their specific functional orientation although they do possess a number of shared objects, some of which are listed below.

1. In each processor, a port connecting the processor to the relay system is used as an object of the computer language. The port is the apparatus and program complex that establishes the processor's communications with the subscriber, creates buffers, and receives and transmits data. In accordance with the protocol for establishing communications, the necessary sequence of messages is developed, relay accuracy is checked, unexpected situations (lack of communications channels, subscriber delays, system mistakes, etc.) are handled and renewed attempts are made to establish connections.

Once the connection is made, data located in the port (buffered) is sent to the subscriber. Meanwhile the receiving port regulates the speed of transmission, identifies and corrects errors and records the received data in the buffer.

2. Another important object of the internal language of these processors is the buffer. A buffer is needed because of the asynchronicity of programs run in different processors and because of the need to gather and temporarily save data that may be called for by a program running in parallel. The data received by the buffer is catalogued and protected. The buffer is a function-adapted object. Depending on the instructions in the WRITE [ZAPOSO] and READ [ChTENIYA] commands, it can present FIFO or LIFO sequences.

3. The efficiency of macropripeline organization of computations is to a large extent determined by the high rate of turnover per input. However, implementation of cyclical programs containing frequent violations of the natural order of commands substantially lowers the actual capacity of the processor involved.

The use of a sequence of numbers, symbols and codes to define the required operations as an object of the processor's internal language is one means of eliminating this discrepancy. Command operands are described by access descriptors, each of which includes a rule for forming a sequence selected from an n-dimensional set, linearly arranged in the memory.

For example, the access descriptor makes it possible to select as a command the row, column or diagonal of a two-dimension set.

4. Objects common to all the processors of a multiprocessor computer are program modules (programs with statically defined data), data sets and data fields.

One process which is a basic object of a multiprocessor computer is its table of access descriptors, each of which determines a call of a program module or data sets, their loading procedure and coordination of communications. Transmission of the process for use in any processor is included within the transmission of the table of access descriptors to that processor.

5. As was already noted, the task is entirely represented by a table of processes whose performance is initiated by the problem control program according to their degree of readiness.

The complexity of controlling parallel processes in solving a problem in a multiprocessor computer makes it necessary to pay special head to the apparatus supporting computer control. For this purpose, a multiprocessor computer has control processors whose internal language, system of objects and object operations are oriented towards programming control tasks. The objects of the control processor's internal language include: apparatus, terminal symbols and their meta-designations, lines of symbols, lists, sequences, stacks, tables, etc.

Operations on objects are implemented by a multilevel system of soft microprogramming which allow the processor to be adapted to run various programming languages efficiency. The following are included to serve as base operations in a microprogramming system: the operations of concatenation and comparison by length and exact coincidence of full ines; division of sublines by coordinates or by membership of its symbols to some group; the operations of assembly and disassembly, location of boundaries and isolation of internal elements of lists. Operations are also provided for work with stacks, sequences, tables, etc.

The use of the interpreted high-level programming language to describe control tasks makes possible the effective interaction of the user and computer and allows the programmer to create and modify distributed operating systems efficiency by joining them with the control programs for the problems being solved (references 5, 6).

Such a technology for programming complex problems makes it possible to put the resources of a multiprocessor computer to the best use, giving the control program a series of specific functions when necessary for checking messages, handling program and external interrupts, analysis of unexpected situations as they arise, etc. In the operating system for a multiprocessor computer, these functions can be implemented in a somewhat generalized form.

Interpretation of the control language makes it possible to dynamically switch unloading modes on and off (without changing the task program), create dynamically dependent objects within the program and change the program without any great loss of time.

The general structure of the task control program includes:

--description of the messages used;
--description of reactions to program interruptions;
--programs for handling external interruptions;
--the actual problem-solving program.

If situations arise that are not described in the problem control program, control is turned over to the next-higher control computer or the computer's operating system.

The system of internal language commands of the control processor, aside from normal operators for assimilation, cycling and branching, contains operators for describing data structures, message exchange, waiting and work with various types of sequences.

The development of languages for control processors and their further sophistication should make it possible to realize V. M. Glushkov's idea of combining within a multiprocessor computer both the final form of a program and its necessary intermediate forms, including those in insufficiently formalized languages, and use these in the task of controlling the computational process. Of course, the uniqueness of such intermediate forms is determined by their subsequent expansion.

If the program is written in high-level languages whose data is not individual words, but entire structures, it is possible in effect to separate parallel processes automatically, plan the organization of their performance, and optimize the composition and distribution of equipment. In this case, it is best to take advantage of the system's adaptability, its capabilities in soft microprogramming and dynamic creation of configurations.

REFERENCES

1. Golovkin, B. A., "Parallelnye vychislitelnye sistemy" [Parallel computer systems], Moscow, NAUKA, 1980, 519 pp.

2. Glushkov, V. M., "Osnovnye arkhitekturnye printsipy povysheniya proizvoditelnosit EVM" [Basic architectural principles for increasing computer capacity], in "Problemy vychislitelnoy tekhniki [Problems in computer equipment], Moscow, MINTI, 1981, pp 6-21.

3. "Organizatsiya vychisleniy v mnogoprotsessornykh vychislitelnykh sistemakh" [Organization of computations in multiprocessor computer systems], Mikhalevich, V. S., Kapitonova, Yu. V., Letichevsky, A. A., et al., in KIBERNETIKA, 1984, No 3, pp 1-10.

4. Glushkov, V. N., Pogrebinsky, S. B., Rabinovich, Z. L., "O razvitiii struktur multiprotsessornykh EVM s interpretatsiey yazyka vysokogo urovnya" [On the development of structures for multiprocessor computers and the interpretation of high-level language], in UPRAVLYAYUSHCHIYE SISTEMY I MASHINY, 1977, No 4, pp 61-65.

5. Pogrebinsky, S. B., "Mashiny dlya inzhenernykh raschetov" [Computers for engineering calculations], in KIBERNETIKA, 1982, No 5, pp 55-60.

6. Pogrebinsky, S. B., Drakh, A. M., Klimenko, V. P., "Apparatnaya podderzhka upravleniya raspredelennymi vychisleniyami" [Instrument support for controlling distributed computations], in "Programmnoye obespecheniye malykh EVM" [Software for small computers], Kiev, IK AN USSR, 1985, pp 3-5.

12261/9835

UDC 681.325.3

Information Code Converter for Control Systems and Data-Processing

18630203b Kiev UPRAVLYAYUSHCHIYE SISTEMY I MASHINY in Russian No 6, Nov-Dec 87 pp 56-60

[Article by A. I. Kondalev: "Information Code Converter for Control Systems and Data-Processing"]

[Text] The present stage of the development of information code converters is characterized by the gerater role played by technological, circuit-engineering, structural and functional methods of increasing system efficiency. All of these methods are inter-related, therefore the desired characteristics --conversion are for the most part achieved through the use of various combinations of the above. Let us examine the possibilities that these methods offer for further improvement of the technical, metrological, economic and performance characteristics of these converters.

The production of systems converters and other computer equipment is carried out in two cycles: the production of the component base and then the addition to that base of the required devices. In their level of integration, many series-manufactured microcircuits are bipolar integrated circuits [IC] and SBIC [not further identified] and are in essence functionally complete assemblies and devices. Nevertheless, they are not a finished product ready for independent operation, but merely components that once combined and coordinated with other components to make items that can operate independently. THe integrated circuits for building systems information code converters are just such a component. They include solid-state and hybrid analogue-digital converters and digital-analogue converters, comparators, analogue multiplexors, analogue signal selection and storage devices, on-line amplifiers, opto-electronic decouplers, etc.

There are in existence scores of technologies for producing bipolar and unipolar (field) analogue and digital (logical) integrated circuits. The most well-known varieties of technologies of bipolar integrated circuits are the TTL [transistor-transistor logic], ESL [not further identified] and I^2L [integrated-injection-logic] circuits while unipolar integrated circuits are represented by p-MOS- [metal-oxide semiconductor], n-MOS-, CMOS- [complementary MOS] and VMOS-structures. These technologies are used

to manufacture both digital and (with some peculiarities) analogue integrated circuits. Regardless of the circuit differences between these IC's, the identical nature of the technology for manufacturing analogue and digital IC's makes it possible to use the same equipment o produce both and also to manufacture both on the same substrates, i.e., to manufacture such analogue-digital bipolar IC's as well as analogue-digital converters, digital-analogue converters and analogue-digital microprocessors, according to their degree of complexity.

In addition, there are specific features in the technology of analogue and digital IC's that are determined by circuit-engineering differences as well asbby the demands placed on both classes of circuits. The chief one is the requirement for precision in stabilization of levels, amplification, conversion and transmission of signals. For analogue IC's, it is necessary to provide a level of precision 2-4 times higher than for digital circuits, i.e., with permissible error levels of 0.1-0.001 percent and plus or minus 10 percent, respectively. Basically, this is the condition that determines the greater demands for precision in the manufacture of analogue IC's.

The deciding factors in the intensive development of the component base for computers are technological and circuit-engineering methods that determine the parameters for improving the characteristics of electronic devices, machines and systems.

The basic parameters of any IC tha are independent of its specific application are its speed, power requirements, reliability and production cost. These parameters depend on the material of the semiconductor substrate, the technology used to manufacture the circuit, the electrical and topological structure of the circuits, the technological design standards, the density of components, degree of integration and the maximum chip area, etc.

For analogue IC's, the speed is the working frequency, while for digital IC's, it is the signal time lag in a single element (gate) of the circuit. The power requirement of analogue IC's is usually the same as for the entire circuit, while in digital IC's, it applies to just that one component. IC reliability is characterized by their resistance to mechanical shock, climatic, electromagnetic and radiational factors. Their production cost includes the cost of materials, equipment and manufacture.

These parameters are all inter-related, therefore, any method of improving one of them influences the others. A more important role is played by the semiconductor material used to build the IC substrate, because this material's physical and technological properties are what determine all other qualities. Thus, the concentration and mobility of electrons and holes in a semiconductor influence the IC's speed, its specific resist determines the power requirements, and the width of the inhibited zone determines the operating and control voltages, current leakage and range of working temperatures, etc.

The best combination of physical and technological properties is found in silicon which lends itself more readily than other semiconductors to purification,

doping, formation of p-n-junctions and insulating layers and can also undergo physical-chemical, thermal, laser, electron-beam and mechanical processing. Silicon forms the basis for the technology of production of bipolar and MOS-transistors used for mass production of numerous varieties of analogue, digital and analogue-digital IC's of various degrees of integration.

The speed of bipolar transistors is limited by many factors, including internal stray capacitances between the base and collector, the base and emitter, collector and substrate, as well as by stray collector, emitter and base resistance. The magnitude of these stray parameters depends on the topology, forms, dimensions and number of electrodes (for example, of emitters), lengths and widths of connecting lines, type of insulation and other geometrical and electrical characteristics. By changing their relation, one can control the manufacture of transistors with increased values in working frequency, amplification factor, power output, identical residual volutage of the base-emitter under different values of residual current, minimal stray capacity between IC components, etc.

Speed is also increased by using transistor-transistor logic structures with Shottki diodes to prevent saturation of bipolar transistors.

ESL-structures, with more complex components than TTL-structures, have the highest speed and power output of all modern bipolar IC's. However, this also creates a problem in removing heat at the higher degree of integration and component density in these circuits.

A distinction needs to be made between other bipolar circuits and I^2L-structures whose dimensions and power requirements have been substantially lowered due to the absence of resistors, but which are also slower than TTL- and ESL-structures.

The unipolar IC technology, which is the basis for the production of MOS-transistors, is a technology that differs greatly from that of bipolar IC's. Just as in the case of bipolar transistors, the speed of MOS-transistors is much influenced by internal stray capacitances and resistances as well as the configuration, dimensions and topology of the source, gate, discharge, and polysilicon and oxide layers.

The speed of MOS-transistors also depends on the type of channel. In n-channels, it is greater than in p-channels since the mobility of the electron-carriers is almost 2.5 times greater than that of hole-carriers, It is true that the technology of n-MOS-structures is somewhat more expensive than that of the p-MOS-structures but the n-MOS-structures have greater relative density and lower power requirements.

The design and structural characteristics of MOS-components make it possible to use them both as transistors and as resistors and capacitors. Their self-insulation dispenses with the need for manufacturing processes to insulate them. The n- and p-channel MOS-transistors can be relatively simply formed in one chip; and this has fostered the development of complementary (CMOS) IC's,

one widely-used variant of MOS-technology. With the reduction in planned sizes to values of 1-2 microns, the dimensions and power requirements of CMOS-structures are now about equal to those of n-MOS structures; but their functional advantages are considerable: very high input resistance, lower power requirement in static mode, high load capacity, good noise resistance and a higher degree of independence from voltage fluctuations, noise and temperature changes. These characteristics of the CMOS-structures are important in the production of high-quality analogue-digital IC's.

In certain situations, to improve one parameter or another in analogue-digital IC's, special MOS technologies are used. Examples of these are double-diffusion (D-MOS), and arrangement of gates in V-shaped grooves (VMOS). The speed of CMOS-structures on a sapphire substrate has at times been increased by reducing stray capacitances between the electrodes and the substrate. Although sapphire costs more than silicon, components made from this material are smaller in size, require less power and are easier to manufacture. This increases the relative density and degree of integration of their microcircuits.

There are many other cost-acceptable technological methods of improving IC parameters, and these methods also offer some potential for further growth. For example, we are far from exhausting the possibilities of reducing circuit sizes, although this is much harder to achieve and much more expensive once we reach values of 0.5-1 micron.

Within each technology, there are broad possibilities for circuit-engineering improvement of the parameters of bipolar IC's and SBIC's of various functional units and devices. The clear confirmation of this are the many varieties of digital-analogue converters that can operate either from standard sources of voltage or current: with one or several (the number of discharges) voltage (current) sources; with decoding matrices on double-weighted resistors or on R-2R; with analogue keys of various types, etc. The structure of a digital-analogue converter is what determines the range of rated values of resistance and current strength, the power requirements, speed, static and dynamic precision, chip size, manufacturing technology and production cost of a bipolar IC. The same can be said of microelectronic analogue-digital converters which exist in an even greater variety. Therefore, it is impossible to choose optimal structure for a digital-analogue converter, analogue-digital converter or other functional device without evaluating the technical and economic characteristics of the corresponding bipolar IC's in the process of their design and development.

An indicator that is often used for comparing digital IC's according to technical quality is the product:

$$\eta_T = t_3 \cdot P \rightarrow \min. \qquad (1)$$

in which η_T is the technical index of IC quality; t_3 is the (mean) time lag of one element (gate), in ns; and P is the power required by the element, in mW.

The lower the product (1), the better the IC. However, this generalized index only partially characterizes IC quality since to it can be added other parameters, such as the degree of integration and IC cost. In this case, the indicator becomes more generalized, technical and economic in nature:

$$\eta_{T\partial} = \frac{t_3 \cdot P \cdot C}{N} \rightarrow \min. \qquad (2)$$

where $\eta_{T\partial}$ is the technical and economic index of IC quality; N is the degree of IC integration and the number of gates in the chip; and C is the cost of making the IC.

One can derive still more general quality indicators by introducing other IC parameters, including functional and structural characteristics like the number of functions performed, number of leads to the frame, etc.

For analogue and analogue-digital IC's, similar indices of quality have been accepted with the difference being that instead of time lag at one rectifier, it is the circuit's working frequency, f_p, that is used and instead of the power required by a single component, the power required by the entire IC, P_c is used. Furthermore, it is more convenient for the numerical values of the indices for better-quality circuits to be greater rather than lower in magnitude. Thus, for the IC's of an analogue-digital converter and digital-analogue converter, the following quality index can be used:

$$\rho_{\partial u} = \frac{f_p \cdot N}{P_c \cdot C} \rightarrow \max. \qquad (3)$$

In the future, the development of parametric methods of improving the quality of digital and analogue IC's will be based on the use of new semiconductor materials, new physical principles, new technologies and new circuit-engineering ideas. The real possibility of increasing the working frequency by as much as 1-10 GHz has been revealed by the production of microprocessor structures from gallium arsenide. The degree of integration of digital IC's already made from this material is close to 10^4 rectifiers per chip while time lag per gate and required power are, respectively, close to 100 ps and 1 mW. An even higher speed is possessed by integrated circuits with Josephson junctions, with a time lag per gate of an average of 25 ps and power of approximately 3 microwatts. However, a great number of complex physical, technological and circuit-engineering problems are yet to be resolved before it becomes economically feasible to mass-product gallium arsenide and Josephson IC's.

A combined technology for bipolar IC's containing bipolar and CMOS circuits within a single chip is highly promising for the creation of high-speed, precision, low-power digital-analogue and analogue-digital converters. An upcoming challenge is the manufacture of bipolar IC's with ultrahigh density, with a reduction of component size by 0.3-0.5 microns through the use of ion-beam and X-ray lithography.

Given the importance of the role played by improving the parameters of the component base, progress in the development of computer technology would not have been as fast if it had not been possible to improve the structure and functions of devices, computers and systems at the same time.

In the case of information code converters, structural methods are based on the use of such functional algorithms and the corresponding converter structures, which have helped to achieve the necessary speed, precision of data conversion and processing, electrical and structural compatibility, operating reliability and economy of production and operation. Owing to the two-cycle production of systems information code converters, the term "structural methods" has been broadened to include two types of structures: circuit-engineering and systems-engineering. The first cycle is associated with the development of circuit-engineering structures and the production of the component base for converters in the form of monolithic IC's for functional i.l. assemblies and devices. The second cycle is associated with the development of systems-engineering structures and the production from IC's of system i.l. converters for specific classes of systems. In some cases, there is an intermediate cycle for development and production of functional i.l. converter assemblies and devices in the form of hybrid IC's that use frameless microcircuits.

Circuit-engineering structures are worked out during the internal synthesis of i.l. converters, when decisions concerning the structural organization of elements, i.l. assemblies and devices are made by optimizing the functional algorithms. The level of circuit-engineering structures is the result of the development of engineering of integrated circuits and we do not separate this from parametric methods within which there is a practically unlimited number of variants for technological and circuit-engineering improvements of converters and computers. A good illustration of this is one of the possibilities of increasing the productivity of converters which results from the creation of monolithic analogue-digital converters with digitalization at an increased beat-frequency, analogue-digital converters with delta-sigma-modulation and analogue-digital converters with interpolated modulation (reference 1). The use of such analogue-digital converters is economical in systems with digital communications, modems and in systems that digitally process high-frequency signals in which the speeds of computations quickly rise.

Systems-engineering structures are worked out at the stage of external synthesis, in which are resolved the problems of optimizing standard sets of systems converters, selection of nomenclature and efficient communications between elements, i.l. assemblies, blocks and devices, and effective interactions between them during the work of the converters in the system. As a whole, the structural enhancing of i.l. converter capacity embraces not only the structure of the converters themselves, but also system structures such as interfaces, communications channels, interrupt chains, memory devices, and coding systems.

The use of structural methods to improve i.l. converters is impossible without considering the technical and economic advantages involved. However, it is hard to evaluate structural efficiency since there is no adequate concept of the "structure" of a parameter or corresponding unit of measure.

Therefore, the structural efficiency is indirectly determined through other indicators. In this sense, formulas 1-3 above to some extent reflect not only the technological, but also the circuit-engineering structural efficiency of IC's. However, this is insufficient for functional devices. Full structural efficiency is a combination of circuit-engineering and systems-engineering structural efficiency and should be determined by the index of quality for performance of the functions for which the device was designed.

The term "performed function" is always content specific: "displacement operation," "addition operation," "multiplication operation," "integration operation" and "analogue-digital conversion," etc. Nevertheless, it is more of a qualitative than a quantitative assessment, for example, cost of manufacture of the device performing the given function (operation) or time needed for the device to perform the assigned function (operation).

Under similar conditions, the structural efficiency of the analogue-digital converters and digital-analogue converters of various structure is well-represented by such indices as speed (time and frequency of conversions) and precision, as well as the more general index of carrying capacity (megabit)

$$r = F_{\text{пр}} \cdot n. \qquad (4)$$

where $F_{\text{пр}}$ is the conversion frequency in MHz; n is the size of the analogue-digital or digital-analogue converter code, in bits.

Since i.l. converters made for the same tasks but having different structures can use IC's of various degrees of integration, indices other than formula 4 are better used to assess the structural efficiency of such converters. One of these is the ratio of carrying capacity to the number of IC units used in the converter:

$$r_k = \frac{r}{m} = \frac{F_{\text{пр}} \cdot n}{m} \to \max,$$

where m is the number of units.

In the case of different types of IC's, the greater the degree of integration, the more useful it is to take the ratio of carrying capacity to the overall cost of IC's used in the converter:

$$r_c = \frac{r}{C} = \frac{F_{\text{пр}} \cdot n}{C} \to \max.$$

The growing demands for expansion of the abilities of i.l. converters to perform computational, logical and control operations in the processing, transmission, storage and use of systems data determine the focus of functional methods for improving their characteristics. The data and program compatibility of these converters and computers, the optimal distribution between them of computational

and control functions, the ability of an i.l. converter to perform operations with analogue and digital data all make it possible to increase system flexibility and productivity substantially, process data in real time and minimize the system's apparatus and software.

The actual output of functional methods will be determined by how closely their development is linked with the development of parametric and structural methods. Systems i.l. converters of various design, analogue-digital processors and other analogue-digital computing equipment made both from bipolar IC's and SBIC's are an integrated combination of the achievements of semiconductor technology, circuit-engineering, systems engineering, algorithmization and programming. Obviously, the development and design of such complex IC's requires some theoretical basis and the use of automation.

The following are the chief factors that must be studied in the development and design of systems i.l. converters to guarantee the optimal resolution of architectural and functional tasks:

--the object of control or study for which the system is created and which is the source and receiver of the data to be converted and processed;

--the computer which does the system's basic data-processing;

--the component base for building the i.l. converter;

--the system requirements for the i.l. converter;

--information about the technical level achieved in the field of i.l. converters.

These factors are sufficient to design system data form converters from a standard component base, i.e. in the second cycle of production of the converters. If customized or semicustomized bipolar IC's are being designed, their manufacturing technology is an additional factor.

The quality and length of the design process depend on the degree to which it has been automated. Therefore, an important cor ition for successfully completing the design process in both cases is : use of CAD, problem-oriented emulation systems and all-purpose com ers for mathematical modelling.

The design of system i.l. converters involves solving the problem of architecture and function, i.e. developing, justifying and selecting the type of architecture and system of functions that are required and sufficient to assure maximum productivity of the control system using the converters.

Assessment of the architectural and functional efficiency and comparison of different variants of architecture at various stages of design are a necessary component of the process of creating a system i.l. converter that meets the demands of T^3 [not furthered identified] (reference 2). In the first stages of designing high-speed i.l. converters, one can use the

rather simple criteria of architectural and functional efficiency:

$$\gamma = \frac{\rho R I n \delta}{C} \to \gamma_{max}.$$

where ρ is the quality index for the bipolar IC's used in the converter (product of the frequency times the degree of integration, related to the magnitude of power requirement); R is the degree of deparallelization of conversion and processing; I is the intelligence index of the data form converter (ratio between the volumes of processing and conversion operations); δ is the precision of data conversion and processing; and γ_{max} is the maximum value of the criteria for architectural and functional efficiency.

The functional of the basic parameters, the optimal value of which corresponds to the system's maximum capacity can serve as an indicator of the system efficiency of the i.l. converter (reference 3).

At the Ukrainian Academy of Science's Computer Science Institute imeni V. M. Glushkov, there has been developed and produced a series of i.l. converters using IC's for control systems, monitoring and automation of scientific research and signals processing. These include the F5286 general-purpose i.l. converter with a built-in fast memory, which is part of the IVK-20 and IVK-6 measurement and computing systems; the i.l. converter of the enhanced-precision analogue-digital converter-36 for the automatic control system for high-precision manufacturing processes and systems of metrological control of measurement instruments; enhanced speed i.l. converters for coding and decoding of brightness and color television signals in experimental digital TV systems; i.l. converters with enhanced noise resistance in the analogue-digital converter-245 for manufacturing process automated control system with a high level general noise; incremental i.l. converters for the analogue-digital converter-112 with a built-in microprocessor designed to code and process signals in real time in systems for automating scientific experiments with high-speed physical processes, and others (reference 3). i.l. Converters are manufactured in instrument or modular versions for mass-produced computers.

As the above material shows, further integrated development of theoretical research and practical design work in the area of parametric, structural and functional methods is necessary for the creation of i.l. converters of the new generation with processor and even computer functions.

REFERENCES

1. Rouz, K. D., "Put' snizheniya stoimosti analogo-tsifrovykh preobrazovateley" [Means of lowering the cost of analogue-digital converters], ELEKTRONIKA, 1986, No 7, pp 57-62.

2. Kondalev, A. I., "Systemna efektyvnist' peretvoryuvachiv formy informatsii" [System efficiency of forms of data-processing], VISTNYK AN URSR, 1984, No 5, pp 43.52.

3. "Vysokoproizvoditelnye preobrazovateli formy informatsii" [High-speed i.l. converters], Kondalev, A. I., Bagatsky, V. A., Romanov, V. A. and Fabrichev, V. A., Kiev, Naukova Dumka, 1987, 280 pp.

12261/9835

UDC 681.325.3

New High-Density Disk Coatings

18630203c Kiev UPRAVLYAYUSHCHIYE SISTEMY I MASHINY in Russian No 6, Nov-Dec 87 pp 71-76

[Article by G. A. Mikhaylov: "New Magnetic Coatings for High-Density Disks"]

[Text] Magnetic disks are an indispensable part of modern computers at all levels, including personal computers. In Soviet computer technology, these disks are distinguished by the fact that they have the weakest production support, in comparison to other computer components. Expansion of the industrial production of disks will most certainly soon become one of the industry's most pressing programs. It will even be necessary within this program to anticipate changes, and this means establishing an orientation to ard production of new types of disks based on the latest scientific research and experimental design. The subject of this article is a brief report on the results of one of these lines of work at the Ukrainian Academy of Science's Institute of Computer Sciences imeni V. M. Glushkov.

The most effective means of improving disks is to increase the density of the data recorded on the surface of the magnetic carrier. This immediately solves three problems: in proportion to the linear density, it increases the speed of data exchange, proportionally to the surface density, it increases data capacity or it reduces disk size.

In order to increase the record density, it is necessary to use new magnetic disk coatings to replace ferrolacquers (reference 1). The latter have a natural density limit on the order of 500-600 shifts in magnetization per millimeter of disk track. It is well known that as the length of a small magnetized section of the track (domain) decreases, its demagnetizing field increases proportionately and somewhat destabilizes the magnetic state. Reducing the thickness of the magnetic surface can restore the best domain proportion and reduce the demagnetizing field, but this also reduces the magnetic flux associated with the domain, i.e., the magnetic signal read by the drive head. Since residual magnetization is low in ferrite materials and the actual magnetized material in ferrolacquer amounts to less than half its mass, these coatings cannot be made any thinner than 0.5-0.6 microns.

There is also anohter obstacle: it is very difficult to produce a thin and homogeneous coating with surface micro-irregularities of less than 0.02 micron when the size of the ferromagnetic particles in the coating cannot be made any smaller than 0.03 micron (otherwise they would become superparamagnetic).

One alternative to a ferrloacquer coating is continuous, thin metallic layers of highly-coercive alloys of cobalt (cobalt-iron, cobalt-nickel, cobalt-chromium, cobalt-nickel-chromium, etc.). Cobalt itself has a high saturation induction and is therefore easily made into an alloy with B_s greater than 1 Tl with a good right-angle hysteresis loop $S = B_r/B_s$ (B_r is residual induction) on a level of 0.8-0.95. Within wide ranges, the coercivity, H_c, of these alloys is, like B_s, regulated to as much as 100 kA/m and higher.

A large reserve of residual induction makes it possible to reduce the thickness of the magnetic coating to 0.03-0.04 microns and to substantially increase recording density. There have been reports of densities of 2,900 shifts/mm (reference 2) using the traditional method of linear recording.

In recent years, efforts have been made to develop and introduce a completely different method of perpendicular recording: domains in the magnetic layer are formed in such a way that their magnetization is directed along the normal to the layer surface and the magnetic properties of the material are such that they allow the vertical domain to remain stable (reference 3). In this case, self-demagnetization does not prevent an increase in recording density, since once the transverse dimensions of the domain are reduced, its demagnetizing field is also weakened. The attainable recording density is estimated at 15,000-17,000 shifts/mm. However, the introduction of this method involves a fundamental reconstruction of all the disk drive apparatus including the magnetic heads, the electromechanical drive and head positioning and the electronics of the write and read channels.

Development of a magnetic disk with perpendicular recording and its mass production are an integral part of the program for development of computer technology, but Soviet scientific research is still not prepared for this task. The beginning stage of this work is the introduction of cobalt coatings for the traditional linear recording. This performs two useful functions: first, it introduces a new magnetic data medium and the technology for its manufacture which practically without any changes, will serve as the stockpile for perpendicular recording; and, second, it greatly improves the electromechanics and the electronics of magnetic disk drives. In particular, disk face pulsing will have to be reduced to 30-40 microns, head positioning precision increased to one micron and the clearance between the disk and floating magnetic head will have to be reduced to 0.15-0.20 microns.

The first stage has achieved linear recording densities of up to 800 shifts/mm and a track density along the disc radius of 40 tracks/mm. At the stage of scientific research, even higher densities were tried without any special problems at recording frequencies of 12 megabits/sec and higher. At surface densities of 800 x 40 = 32,000 shifts/mm^2, the data capacity of normal-sized

disks (nonformatted recording on both sides) will be:

Disk diameter (mm)	Disk data capacity (megabytes)
356	750
203	240
133	100
89	46

Such data capacities for individual disks are 1.5-2 times higher than the capacity of disk drives with permanent disks, which have been named the fourth phase in the program of upcoming work on the concept of SM computers [not further identified]. As metal-coated disks come into greater use, these capacities can be increased 2-3 times, since recording densities of 86,000 bits/mm^2 have already been tested on this coating (reference 2).

To apply thin metal coatings to disks, two technologies have been developed in parallel. They are electrolytic precipitation and vacuum condensation. In this case, one variant of vacuum condensation was chosen: evaporation of initial components of the alloy from a ceramic crucible heated by a molybdenum helix and condensation of the coating through a diaphragm onto the surface of a rotating disk in a chamber with a working vacuum level of 10^{-5} torr. Vacuum condensation makes it possible to use a wider selection of starting materials for the coating, more easily assure their composition and produce thinner, denser and purer coatings free of chemically aggressive components that weaken the corrosion-resistance of the disk.

To prepare an experimental-industrial technology for applying magnetic disk coatings, it is necessary to clarify and describe a complex chain of relationships starting with the required working characteristics of the coating and ending with the initial technological parameters. Between these two end links several intermediate ones, all of which are set and defined by technology. For example, if we look at working characteristics such as amplitude of the reproduced signal, the signal-noise ratio, the maximum recording density, etc., then it is necessary at the start to determine their relationship to magnetic characteristics such as coercivity, residual magnetization, right-angle hysteresis loop, etc. Furthermore, it is also necessary to know the relationship of magnetic properties to thickness, structure and microstructure of the coating, and finally, to know the relationship of these properties to the initial composition of the alloy, temperature and velocity conditions, etc. No less complex a matter is the indirect connection between technological conditions and such coating properties as abrasion-resistance, surface quality, etc. This stage has involved very extensive scientific research especially the experimentation part, which is briefly described below.

Let us turn to the characteristics reproduce the signal and their relationship to the magnetic properties of the coating.

The signal amplitude, E, is proportional to the value of the shifted magnetic flux, associated with the memory element, and to the product $B_r \cdot T$, residual

induction ($B_r = \mu_0 M_r$, where M_r is residual magnetization) per thickness of the magnetized layer; signal amplitude is inversely proportional to the length of the shift, i.e., the width, a_t, of the transition section dividing adjacent memory elements--domains with opposite magnetization. Figures 1 and 2 show the nature of this relationship. The width of the transition depends both on the thickness, t, and on the coercivity, H_c. At first estimate, a_t is proportional to $M_r \times t/H_c^{1}$, and therefore, to some extent at least, an increase in H_c contributes to an increase in signal amplitude. Sometimes, the relationship of E to M_r and H_c is evaluated by the empirical formula $E \sim M_r^n \times H_c^m$, where at high recording densities, m is on the order of 0.9, and n is around 0.1 (reference 4).

The signal-noise ratio for these coatings is usually around 36-37 dB (reference 2); however, it deteriorates as recording density rises (reference 5).

As recording density, D, rises, the amplitude of the reproduced signal quickly drops, since the effects of the demagnetizing field of the domain grow progressively stronger as its length is diminished. From this perspective, coating quality is characterized by the relationship of the signal to the spatial frequency (see figure 2). The point D_{50} in the relationship is usually indicated; this is the point at which the signal has decreased to 50 percent of its original value, i.e., it has reached a level of 6 dB. Sometimes the point D_{70} is used as the limiting value for density, instead of D_{50}. D_{70} is at 3 dB, or 70 percent of the "low-frequency" amplitude. The characteristics E(D) of thin-film metallic coatings are in the realm of higher densities than those attainable with ferrolacquer coatings. They are basically determined by the relationship between the demagnetizing field of the domain and the coercivity. The previously mentioned relationship $M_r \times t/H_c$ gives a good assessment of the minimum length of the domain resistant to demagnetization and, consequently, the maximum recording density (reference 6).

One can raise the maximum density while reducing the coating thickness to the limits established by an acceptable value for signal amplitude and good technological conditions (even thickness, reproducibility and homogeneity of properties). One can raise density while increasing coercivity, not only by reducing thickness, but by other methods described below. This process is illustrated in figure 3, and its limitations are the high currents of recording and high saturation magnetization of the material of the ferrite recording head. With the transition to a new high-coercivity carrier, it is also necessary to switch to magnetic heads made of MnZn ferrite that has a magnetization saturation on the order of 400 kA/m, which is approximately 1.4 times higher than NiZn ferrite.

Apparently, the attainable coercivity values will be on the order of 70-80 kA/m.

The full potential of a magnetic carrier can be reached only in a "magnetic head-disk" system of adequate quality. If the gradient of the recording head field penetrating the magnetic layer does not provide a narrow transition, then

the recording quality will drop. There should also be a sufficiently-strong connection between the magnetic field of the domain and the head when it is reading. Therefore, in using thin-film metallic disk coatings, the heads work with diminished height, d, of head floatation over the disk and length, g, of the gap between the poles at the tips of the head. The floating height, d, greatly affects the value of signal amplitude loss due to inefficient recording and reading. They are usually measured as kd/λ (dB) where k is a constant on the order of 30-90 and $\lambda = 1/D$, the recording "wavelength".

In analytical models for reading heads, there is usually a multiple of the type $(-2\pi(a_t+d)/\lambda)$, which expresses the weakening of the signal as the recording density, head-disk clearance and width of transition are increased (references 4, 7). To obtain high recording densities on thin-film coatings, heads with a gap length of g = 0.20-0.46 microns at a floating height of d = 0.12-0.22 microns were used (references 2, 8, 9).

Since the working characteristics of the coating are determined by its magnetic properties, it is important to know what influences the technology chosen can have on these properties.

The magnetic properties of a coating are above all determined by cobalt, the basic alloy component (about 80 percent). Pure cobalt has a high induction saturation ($B_s = 1.79$ Tl) and magnetization ($M_s = 1.14 \times 10^5$ A/m), but low coercivity ($H_c = 2\text{-}4$ kA/m). The purpose of adding to the cobalt other magnetic (iron, nickel) or nonmagnetic (chromium, tungsten, phosphorus, oxygen, etc.) components is first of all to lower somewhat ("dilute") its high magnetization and second, to increase substantially its coercivity by changing the alloy's crystalline structure and microstructure.

Other methods are also used to change the structure and microstructure. for example, cobalt is crystallized on a chromium sublayer at 200-300°C when its crystalline lattice has a hexagonal form, this radically changes its magnetic properties and in particular raises its coercivity by several factors. By manipulating the proportions of alloy components, thickness of the magnetic layer and chromium sublayer, substrate temperature during condensation and speed of condensation, while selecting a range of sufficiently even relationships, one can produce coatings with the desired reproducible properties.

Figures 4-2 present graphs from references 1 and 10-13 which show possibilities for controlling the magnetic properties of coatings. Figure 4 introduces the relationships for coercivity of cobalt film, which show the influence of magnetic layer thickness, thickness of the chromium sublayer and temperature of the substrate, respectively. For high-density magnetic recording, the coercivity of a cobalt film is insufficient, and its residual magnetization is too high. It is therefore better to use cobalt-chromium, cobalt-nickel, etc. alloys. Nonmagnetic chromium lowers magnetization in proportion to its percentage of the alloy; and at 23-25 percent chromium, the alloy becomes nonmagnetic (see figure 5). In this case, the coercivity at first rises and then sharply drops (reference 10). The coercivity of coatings made from cobalt-chromium alloy at a thickness of less than 0.5 microns and applied

without a chome sublayer is too low (28-30 kA/m) for high-density linear recording. The maximum coercivity with low magnetization is reached at a level of 18.5 percent chromium. Furthermore, the influence of chromium content on the magnetic properties of this alloy is substantial, and since the volatility of the alloy's contents varies, there are difficulties in its manufacture.

The introduction of a chromium sublayer, which contributes to the formation of a high anisotropic hexagonal phase of the alloy, substantially improves the situation by creating as much additional coercivity as is needed and flattening out the relationship between coercivity and chromium content (see figure 6) (reference 11). An essential concern here is to prevent oxidation of the sublayer surface before the basic layer is deposited. This layer must be deposited immediately after the sublayer is laid (references 12, 13).

For technological reasons and for convenience, similar and perhaps even better (for linear recording) coating characteristics are obtained with cobalt-nickel alloy on a chromium sublayer. Magnetic nickel (B_s = 0.6 Tl, M_s = 0.387 x 10^5 A/m) does not strongly lower the magnetization of the alloy--with 25 percent nickel, the magnetization is only 17 percent lower than in pure cobalt. Coercivity of the alloy is high. Figure 7 shows that it is 20-22 A/m higher in a cobalt-nickel coating than in a cobalt coating (both of them on a chrome sublayer) in all thickness ranges from 0.02-0.1 microns.

Study of the magnetic properties of coatings and their connection with the structure and manufacturing parameters show that out of the various combinations of cobalt, cobalt-chromium or cobalt-nickel, the best characteristics are displayed by coatings made from coablt-chrome and cobalt-nickel alloys on a chrome sublayer. They make it possible to select the required magnetization and high coercivity within a range of fairly smooth relationships to technological factors, such as substrate temperature, condensation speeds, quality of the vacuum in the chamber, etc. Furthermore, the great difference in alloy component evaporation speeds (by a factor of three or more) and the strong relationship between coercivity thickness to the magnetic layer makes it necessary to maintain correct manufacturing conditions very strictly.

There are also available a number of other means of controlling magnetic properties. These have been studied in some detail and successfully used. One can raise coercivity by thermal annealing of the magnetic coating, namely, keeping it under high temperatures (250-350°C) for a period of several hours. Coercivity is raised or anisotropic properties are gained if during condensation of the coating of the stream of atoms from the evaporating alloy strikes the substrate at an angle, rather than perpendicular to it, thus forming an elongated crystalline structure. Literature on the subject offers other methods, such as preliminary structuring of the disk surface, creation of contours with controlled surface unevenness, etc.

Along with the basic technology, there have also been developed other methods for obtaining the required mechanical and strength properties. These include strengthening the surfaces of disks made from aluminum alloys

that are not hard enough, and application of a protective coating on top of the magnetic layer.

Reinforcement coatings made from nickel-phosphorus or nickel tin alloys used in electrolytic precipitation onto aluminum disks have been proven impractical for vacuum condensation. Following manufacturing processes under temperatures over 200°, they introduce distortions in the mangetic characteristics of the basic coating. To some extent, chromium sublayers have helped strengthen aluminum bases. However, the best method for vacuum technology is apparently magnetron spraying of a thin sublayer of powdered metal, such as stainless steel (reference 14).

Vitreous and devitrified glass disks are very promising for vacuum condensation because they have the required hardness and their coefficients of thermal expansion match those of the alloys more closely.

Thin-film metallic magnetic coatings need to be top-protected by a very thin (0.06-0.08 microns) and strong additional layer. Such protection is not required for a ferrolacquer coating whose binding component has mechanical properties that assure its nominal wear during operation. In many ways, a protective coating determines the conditions under which the head and disk work together. It should be very thin so it will not be necessary to increase the distance between the head and magnetic layer and change the recording density; roughness and friction coefficients are limited to assure normal floating of the head over the disk and prevent sticking; the hardness of the protective layer hsould guarantee the coatings resistance to wear.

Of the many tried protective compositions (titanium nitride, boron-palladium and boron-chromium compounds, the natural oxide of the magnetic layer, etc.), the best results have been achieved with silicon dioxide (references 15, 16). The international standard for wear-resistance stipulates that the amplitude of the read signal not be lowered more than 10 percent after 20,000 cycles of disk starts and stops accompanied by head contact with the disk. Protective coatings made from silicon dioxide of average quality (4-percent vitreous) on vitreous disks have shown that the signal is reduced by only 1.8 percent. Somewhat poorer results were obtained with metallic disks in which the signal amplitude dropped 4.7 percent after 10,000 cycles.

Study of the properties of new magnetic coatings for disks with a high recording density, development of technology for application of these coatings and testing of experimental coatings in mock-up disks have provided extensive and reliable information about the chief aspects of this problem. This information can be used to make the responsible decision to give greater emphasis to the development of this technology and to organize experimental industrial production of new disks.

It has been established that high-coercivity magnetic disk coatings (thin layers of cobalt-nickel or cobalt-chrome alloys on a chrome layer) have a substantially higher recording density than ferrolacquer coatings and can therefore be used to create more compact disks with a greater data capacity and greater speed. With regard to manufacturing plans, the new coatings are bringing us closer to a more progressive method of magnetic recording--perpendicular recording with higher density.

The technology for application of coatings by condensation from the vapor phase in a vacuum is sufficiently controllable, stable and reproducible. Although vacuum condensation is inconvenient for organization of a continuous manufacturing process, the capacity of vacuum systems can be increased by the addition of built-in multipositional manipulators within the vacuum chamber and use of the abundant arsenal of equipment offered by modern microelectronics.

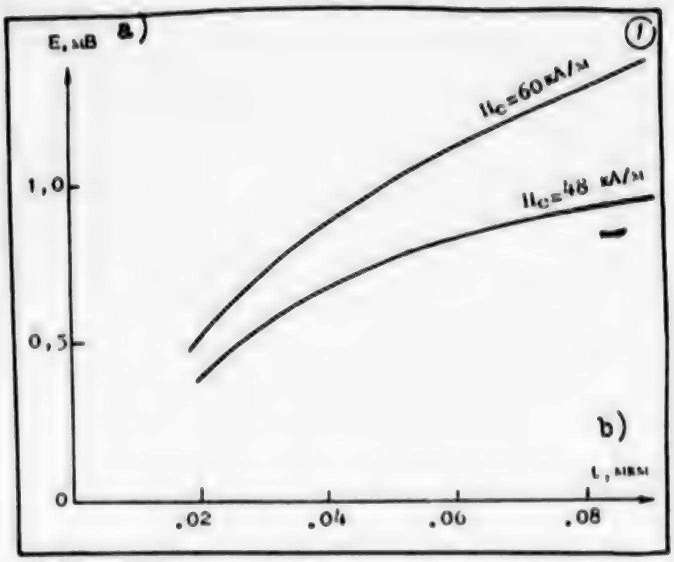
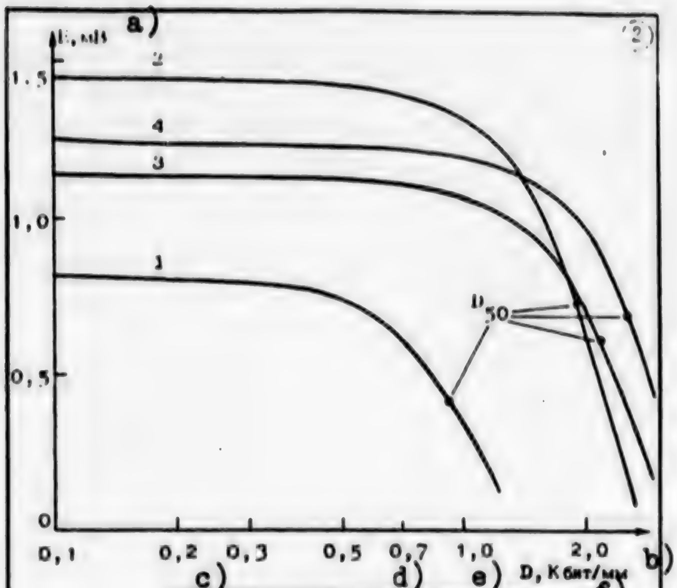


Figure 1. Relationship of amplitude, E, of a reproduced signal to thickness, t, of the magnetic coating and its coercivity, H_c (at a recording density of D = 600 shifts/mm).

Key:
- a) mV
- b) microns



Кривые	H_c, кА/м	t, мкм	g, мкм	d, мкм
1	56	0,17	0,80	0,27
2	56	0,17	0,35	0,11
3	56	0,12	0,35	0,11
4	100	0,12	0,35	0,11

Figure 2. Relationship of amplitude, E, to recording density, D, at different values of H_c, t, g and d (reference 9).

Key:

a) E,mV
b) D,Kbits/mm
c) curve
d) t, microns
e) g, microns
f) d, microns

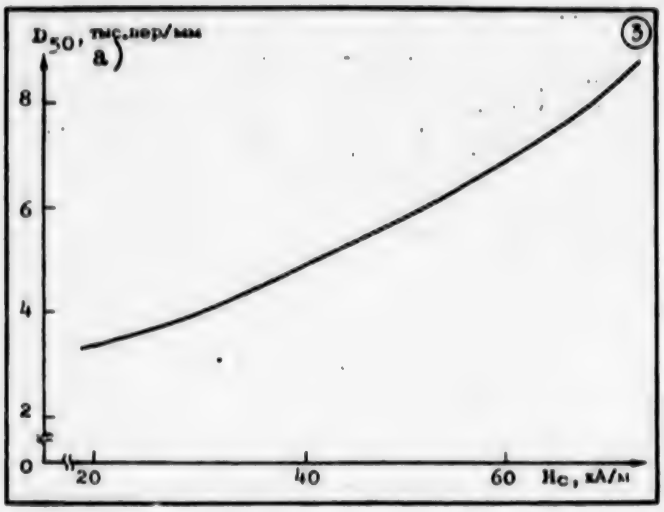


Figure 3. Relationship of recording density limit D_{50} to strength of coercivity, H_c, of the coating at a coating thickness of 0.1 microns.

Key:
a) thousands of shifts/mm

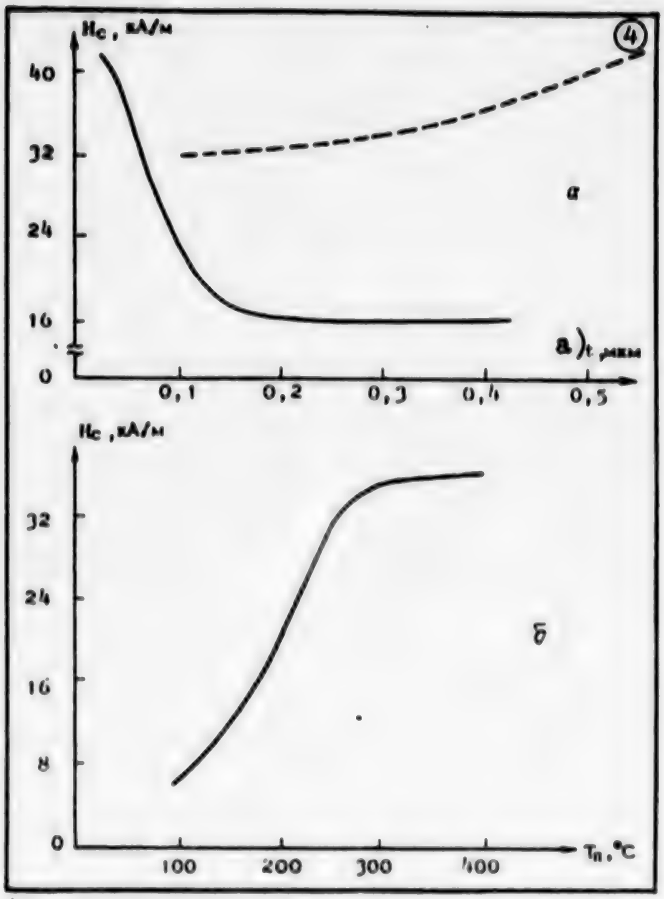


Figure 4. Relationship of strength of coercivity, H_c, of cobalt films on a chromium sublayer: a--to thickness, t(Co), (continuous line) and to the thickness of the chrome sublayer, t(Cr), at t(Co) = 0.04 microns (dotted line); b--to substrate temperature, T_π.

Key:
a) microns

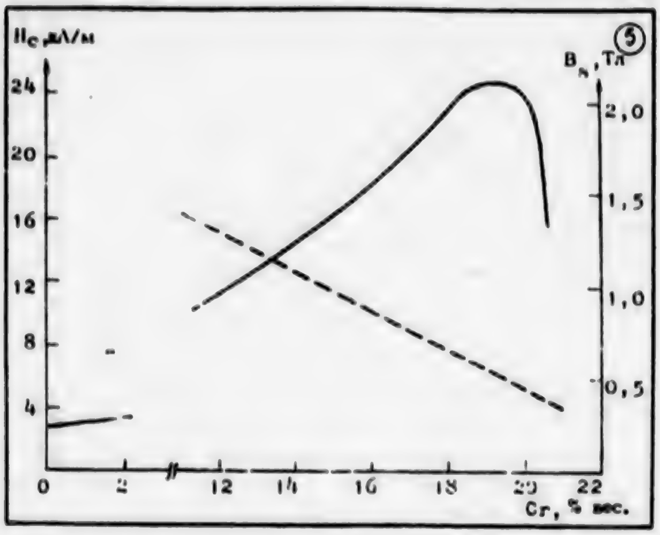


Figure 5. Relationship of strength of coercivity, H_c, (continuous line) and induction saturation, B_s, (dotted line) of cobalt-chromium coatings to chromium content (thickness range of 0.12-0.3 microns).

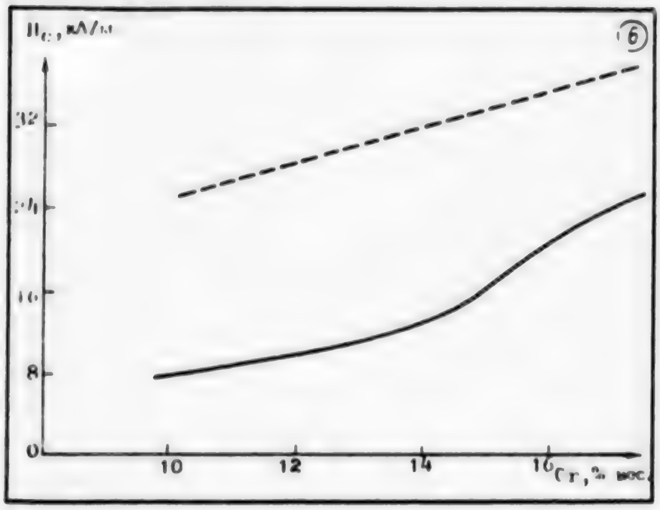


Figure 6. Relationship of strength of coercivity, H_c, of cobalt-chromium films on glass substrates (continuous line) and on a chromium sublayer (dotted line) to chromium content in films.

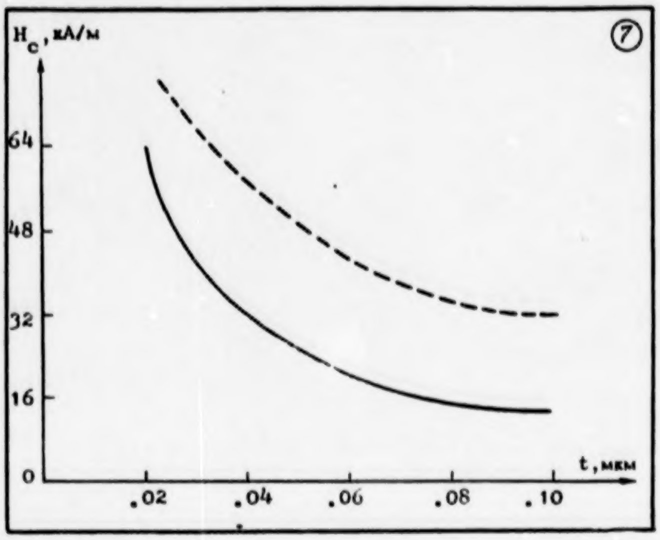


Figure 7. Relationship of strength of coercivity, H_c, of cobalt (continuous line) coatings and an alloy of 75 percent cobalt and 25 percent nickel (dotted line) to coating thickness, t. The chromium sublayer thickness is 0.3-0.4 microns.

REFERENCES

1. Mikhaylov, G. A., Zharikov, G. P., Moroz, A. A., "Sovremennoye sostoyaniye razrabotok magnitnykh pokrytiy dlya diskov pamyati s vysokoy plotnostyu zapisi informatsii" [Current status of the development of magnetic coatings for memory disks with a high data-recording density], in "Sovremennye vidy nositeley dlya vosproizvedeniya, khraneniya i raspredeleniya informatsii: Doklad Vsesoyuznoy nauchno-tekhnicheskoy konferentsii" [Modern types of carriers for reproduction, storage and distribution of data: Report to the All-Soviet scientific and technical conference], Yerevan, 1981, pp 68-85.

2. Suganuma, Y., Tanaka, H., "Production Process and High Density Recording Characteristics of Plated Disk," IEEE TRANS. ON MAGN., 1982, Vol 18, No 6, pp 1215-1217.

3. Karpenkov, S. Kh., "Problemy i perspektivy perpendikulyarnoy magnitnoy zapisi" [Problems and prospects for perpendicular magnetic recording], ZARUBEZHNAYA RADIOELEKTRONIKA, 1984, No 6, pp 56-71.

4. Koester, E., Pfefferkorn, D., "The Effect of Remanence and Coercivity on Short Wavelength Recording," IEEE TRANS. ON MAGN., 1980, Vol 16, No 1, pp 56-58.

5. Tong, H., Ferrier, R., "The Micromagnetics of Thin-Film Disk Recording Track," IEEE TRANS. ON MAGN., 1984, Vol 20, No 5, pp 1831-1833.

6. Khan, M., Lee, J., "Comparison of Plated Co-P and Sputtered Co-Re As Recording Media Based on Nucleating Growth, Structure and Magnetics," JOURNAL OF APPLIED PHYSICS, 1985, Vol 57, No 1, pp 4028-4030.

7. Kigya, F., "Studies of Spacing Loss in Longitudinal and Perpendicular Recording," JOURNAL OF APPLIED PHYSICS, 1984, Vol 55, No 6, pp 2220-2222.

8. Wood, R., "An Experimental Eight-Inch Disk Drive wtih One Hundred Megabytes per Surface," IEEE TRANS. ON MAGN., Vol 20, No 5, pp 698-700.

9. Kaneko, R., Koshimoto, Y., "Technology in Compact and High Recording Density Disk Storage," IEEE TRANS. ON MAGN., 1982, Vol 18, No 6, pp 1221-1223.

10. Zharikov, G. P., Gorbachev, V. V., "Magnitnye svoystva vysokokoertsitivnykh pokrytiy dlya diskov" [Magnetic properties of high-coercivity coatings for disks], in "Zapominayushchie ustroystva i krioelektronnye komponenty EVM" [Memory devices and cryoelectronic components of computers], Kiev, IK AN USSR, 1978, pp 47-51.

11. "Kondensirovanniye v vakuume pokrytiya iz magnitnykh splavov dlya diskov vneshney pamyati s bolshoy plotnostyu zapisi" [Vacuum condensation of a magnetic-alloy coating for external memory disks with high recording density], Zharikov, G. P., Gorbachev, V. V., Zorin, V. V., Moroz, A. A., in "Fiziko-tekhnologicheskiye osnovy integralnykh elementov EVM" [Physical and technological bases of integral computer components], Kiev, IK AN USSR, 1981, pp 38-46.

12. Zharikov, G. P., Zorin, V. V., Moroz, A. A., "K metodike kondensatsii magnitnykh pokrytiy iz sloyev khroma i kobalta dlya diskov vneshney pamyati" [Towards a methodology of condensing magnetic coatings from chromium and cobalt layers for external memory disks], in "Zapominayushchiye Ustroystva" [Memory devices], Kiev, IK AN USSR, 1980, pp 53-57.

13. "O vliyanii protsessov okisleniya pri kondensatsii na koertsitivnuyu silu magnitozhestkikh pokrytiy iz sloyev khroma i kobalta" [On the effects of oxidation processes during condensation on the coercivity of fixed-magnetic coatings from layers of chromium and cobalt], Zharikov, G. P., Zorin, V. V., Mikhaylov, G. A., Moroz, A. A., in "Novye magnitnye materialy dlya mikroelektroniki: Tez. dokl. Vsesoyuz. shk-seminara" [New magnetic materials for microelectronics: Abstracts from reports to the All-Union school seminar], Ashkhabad, 1980, pp 47-49.

14. Rossi, E., McDonough, G., "Vacuum-Deposited Thin-Metal Film Disk," JOURNAL OF APPLIED PHYSICS, 1984, Vol 55, No 6, pp 2254-2256.

15. Oleynik, L. I., Zorin, V. V., "K voprosu o vakuumnoy tekhnologii izgotovleniya zashchitnykh pokrytiy diskov vneshney pamyati" [On the problem of vacuum technology for manufacturing protective coatings on external memory disks], in "Fiziko-tekhnologicheskiye voprosy kibernetiki" [Physicotechnological problems of computer science], Kiev, IK AN USSR, 1983, pp 57-65.

16. Oleynik, L. I., Zorin, V. V., "Issledovaniye nekotorykh mehanicheskikh svoystv kvartsevogo zashchitnogo pokrytiya magnitnykh diskov" [Study of some mechanical properties of quartz protective coatings for magnetic disks], in "Fiziko-tekhnologicheskiye voprosy kibernetiki" [Physicotechnological problems of computer science], Kiev, IK AN USSR, 1985, pp 65-70.

12261/9835

- END -

END OF

FICHE

DATE FILMED

31 Oct 88